30453

Topology for Efficient Information Dissemination in Ad-hoc Networking

Esther Jennings
jennings@arcadia.jpl.nasa.gov
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109

C. M. Okino
cokino@dartmouth.edu
Thayer School of Engineering
Dartmouth College
Hanover, NH 03755

Abstract—In this paper, we explore the information dissemination problem in ad-hoc wireless networks. First, we analyze the probability of successful broadcast, assuming: the nodes are uniformly distributed, the available area has a lower bound relative to the total number of nodes, and there *is zero knowledge of the overall topology of the network. By showing that* the probability of such events is small, we are motivated to extract good graph topologies to minimize the overall transmissions. Three algorithms are used to generate topologies of the network with guaranteed connectivity. These are the minimum radius graph, the relative neighborhood graph and the minimum spanning tree. Our simulation shows that the relative neighborhood graph has certain good graph properties, which makes it suitable for efficient information dissemination.

I. Introduction

The concept of collectively utilizing distributed sensor modules in a hierarchical manner was first introduced as *cooperative sensor networking* [1]. An extension of this idea is to prolong the life-time of finite energy sources by leveraging *cooperative modulation techniques* [2]. This technique relies heavily on the efficient usage of battery power on the local communication link and requires some sharing of information, which motivates our investigation into the information dissemination problem.

The connectivity among nodes directly influences the efficiency of information dissemination within a network. Conventionally, the topology of an ad-hoc network is defined by the transmission radius d of each node. Due to the dynamic and ad-hoc nature of such networks, using a fixed d might not render a connected network at all times. Sometimes, the network is partitioned into several connected components where each component is a connected sub-network, but there are no connections between the different sub-networks; we call this a *partitioned* network.

In [3], Gupta and Kumar showed that, given n nodes such that each node covers a RF circular area $\pi d_{RF}^2 = \frac{\log n + c(n)}{n}$, then the network approaches connectivity with probability 1 as $c(n)$, the connectivity measure in [3] approaches infinity, synonymous to the number of nodes approaching infinity. We examine the alternate extreme of relatively sparsely spaced nodes and the probability of distributing a piece of information in a multi-hop manner from a node to all other nodes in the network, with a fixed number of transmissions.

In this paper, we examine information sharing (as *gossip*) for the objective of leveraging cooperative modulation techniques that requires each node to communicate its information to all other nodes. Formally, the *gossip* problem is defined in [4] as the all-to-all communication problem, where each node holds a piece of independent (or disjoint) information. The set containing the union of information from all nodes is called the *cumulative* information of G. The gossip problem is to find a communication strategy such that each node from the set of all nodes, V acquires the whole cumulative information.

In the cooperative sensor network model, aggregation and multiplexing might not result in energy-per-bit gains. Thus, the gossip problem can be considered as n separate broadcasts, where n is the number of nodes. We analyze the probability of successful broadcasting from a source node, using a fixed number of transmissions. We assume the nodes to be randomly placed, and there is no knowledge of the topology of the network. As expected, for large areas (i.e. areas where the denseness of the nodes is not a considerable factor), the probability of successful broadcast is low, motivating our investigation into specific network graph topologies guaranteeing connectivity for information dissemination. We propose three classes of graphs and examine their graph properties to determine their suitability for broadcasts in a gossip manner.

The research described in this publication was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Dr. Okino was supported by a NASA Summer Faculty Fellowship.

II. Probability Analysis of Blind Broadcast

We now characterize the total number of transmissions required for blind broadcast, which is topology independent and without any knowledge of the topology. The node originating the broadcast is the *source node*. We now formally define connectivity.

Let $l_i, l_j \in \mathcal{R}^2$ be the locations of nodes v_i and v_j respectively, where $v_i \neq v_j$. Direct connectivity between any pair of nodes v_i and v_j is defined by the transmission radius d. Specifically, for v_i and v_j, we have $\| l_i - l_j \| \leq d$, where the norm used is the Euclidean norm (i.e., L^2-norm). We say that v_i and v_j have multi-hop connectivity if there is a non-empty set of nodes P with size $|P|$, where the nodes of P are labeled $a(1), a(2), \ldots, a(|P|)$, and we have $\| l_i - l_{a(1)} \| \leq d$; for $2 \leq k \leq |P| - 1$, we have $\| l_{a(k)} - l_{a(k+1)} \| \leq d$; and finally $\| l_{a(|P|)} - l_j \| \leq d$.

We say that a set of nodes are *connected* if each pair of nodes is either directly or multi-hop connected. Otherwise, the set of nodes is *partitioned*.

Let $l_{src}, l_1, l_2, \ldots, l_{n-1} \in \mathcal{R}^2$ be the locations of the source node and nodes $v_1, v_2, \ldots, v_{n-1}$ respectively, and let V contain all the nodes. Let $\mathcal{N}(l_i)$ be the maximal set of nodes contained in the area of radius d centered at l_i. Specifically, we have $\mathcal{N}(l_i) = \{v_z : v_z \in V \text{ and } \|l_i - l_z\| \leq d\}$.

Assume each node has a transmission radius of d, the area covered by a node v_i is πd^2 centered at l_i. We use $A(v_i)$ to denote the radial area covered by v_i. Suppose nodes x and y are directly connected. Ignoring edge effects, the maximum area where a third node z can reside such that z is directly connected to y but not to x, is upper bounded by $A(y) - [A(x) \cap A(y)]$. Let $\alpha = 2\pi + 3\sqrt{3}$.

Lemma 1: Suppose there exists nodes v_i, v_j at location $l_i, l_j \in \mathcal{R}^2$ respectively, such that $\|l_i - l_j\| = d$. If each node can cover an area with radius d, the non-overlapping area of either node v_i or v_j is $\frac{\alpha d^2}{6}$.

Proof of Lemma 1 : Consider Figure 1, where two nodes are separated by distance d, and RF radial transmission distance of nodes v_i and v_j are d_i and d_j respectively, where $d_i = d_j = d$.

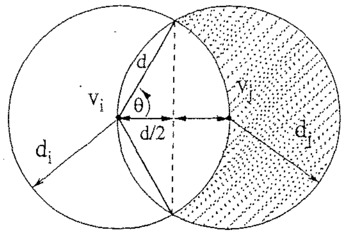

Fig. 1. Non-overlapping area with respect to Node v_j.

Each node covers area πd^2. Since l_i (Node v_i) and l_j (Node v_j) are exactly d apart, then there exists a perpendicular line between l_i and l_j such that it bisects the line joining l_i and l_j. Thus, we can compute θ in Figure 1 as $\theta = \cos^{-1}(\frac{d/2}{d}) = \cos^{-1}(1/2) = \pi/3$. Since the angle representing the overlap is twice the angle θ, the area encompassed by the arc of the two points is $\frac{1}{2}2\theta d^2 = \frac{\pi d^2}{3}$.

Subtracting the triangular area encompassed by the arc, we have $\Delta = \frac{\pi d^2}{3} - 2 \cdot \frac{1}{2}(\sqrt{3}\frac{d}{2})(\frac{d}{2}) = \frac{\pi d^2}{3} - \frac{\sqrt{3}d^2}{4} = \frac{(4\pi - 3\sqrt{3})d^2}{12}$.

Since the overlap occurs on both sides of the perpendicular line bisecting the line joining l_i and l_j, we have $2\Delta = \frac{(4\pi - 3\sqrt{3})d^2}{6}$. To obtain the amount of area $A_{non}(d)$ that is not overlapping, we have

$$
\begin{aligned}
A_{non}(d) &= \pi d^2 - 2\Delta \\
&= \pi d^2 - \frac{(4\pi - 3\sqrt{3})d^2}{6} \\
&= \frac{(2\pi + 3\sqrt{3})d^2}{6} .
\end{aligned}
$$

□

For a number of randomly placed nodes each having a transmission radius d, we can write a topology connectivity probability in terms of the exact number of transmissions required to propagate a bit of information to all other nodes.

Let T_{src} be the number of transmissions required to broadcast error free in a multi-hop manner from the source node. Assume that all the nodes n reside in an area A_n such that

$$
A_n \geq \frac{6\pi + \alpha(n-2)}{6} d^2 . \tag{1}
$$

Theorem 2: Consider n nodes with broadcast radius d randomly placed over area A_n with uniform distribution. The upper bound on the probability of a node requiring $T \doteq k$ transmissions to propagate a bit of information to all other $n-1$ nodes is upper bounded by $G(k,n)(d^2/A_n)^{n-1}$, where

$$
G(k,n) = \frac{6\pi(2\pi + 3\sqrt{3})^{k-1}[4\pi - 3\sqrt{3} + \alpha k]^{n-1-k}}{6^{n-1}} .
$$

Using (1), we have

$$
Pr\{T_{src}(n) = k\} \leq \frac{6\pi\alpha^{k-1}[4\pi - 3\sqrt{3} + \alpha k]^{n-1-k}}{(6\pi + \alpha(n-2))^{n-1}} .
$$

Proof of Theorem 2:

Let $l(src)$ be the location of the source node. Let l_i be the location of v_i, where $1 \leq i \leq n-1$. For the broadcast of a bit from the source requiring a single transmission to another node v_i, the receiving node must be within radius distance d of the source node, and so the probability of a single transmission to a single node, $Pr\{T_{src}(1) = 1\}$, we have

$$
\begin{aligned}
Pr\{T_{src}(1) = 1\} &= Pr\{l_i \in A(l_{src})\} \\
&= \frac{\pi d^2}{A_n} \triangleq p_{first} .
\end{aligned} \tag{2}
$$

The probability of a single transmission from the source node to all other $n-1$ nodes is p_{first}^{n-1}. For nodes to require multiple transmissions, the spatial allowable area not in contact with any other nodes, and thus requiring more transmissions, is upper bounded by the non-overlapping area between two nodes w.r.t. one of the nodes. Thus, the probability of broadcasting a bit over any transmission other than the first transmission, p_{trans} can be written as the probability of the source node propagating through v_i to transmit a bit to v_j such that v_j is at a distance larger than d. The number of transmissions is lower bounded by the number of hops. Thus, using Lemma 1, we have

$$
\begin{aligned}
p_{trans} &\leq Pr\{l_i \in \mathcal{N}(l(src)) \bigcap l_j \in \mathcal{N}(l_i)/l_j \notin \mathcal{N}(l(src))\} \\
&= \frac{(2\pi + 3\sqrt{3})d^2}{6A_n} \\
&= \frac{\alpha d^2}{6A_n} .
\end{aligned} \tag{3}
$$

For a total of n nodes where the source node needs to propagate over i transmissions to all of the other $n-1$ nodes, there are $n-1-i$ nodes that are allowed to placed anywhere within the allowable area of transmission. Thus, the probability of the $n-1-i$ nodes residing within the transmission area has an upper bound of

$$
\begin{aligned}
p_{others}(i) &= \frac{1}{A_n}\left(\frac{6\pi d^2 + (2\pi + 3\sqrt{3})d^2(i-1)}{6}\right) \\
&= \frac{[4\pi + 2\pi i + 3\sqrt{3}i - 3\sqrt{3}]d^2}{6A_n} \\
&= \frac{[4\pi - 3\sqrt{3} + \alpha i]d^2}{6A_n} .
\end{aligned} \tag{4}
$$

Using (2), (3), and (4), we can upper bound the legitimate area required in order to transmit $T_{src}(n) = k$ as the probability of at least one node contained in the initial space, at least one node

30453

contained in each of the non-overlapping spaces (equivalent to the $k-1$ transmissions), and all other nodes contained anywhere among the allowable space. Thus, we have

$$\begin{aligned} Pr\{T_{src}(n) = k\} &\le p_{first} \cdot p_{trans}^{k-1} \cdot p_{others}(k)^{n-1-k} \\ &= \frac{\pi d^2}{A_n} \cdot \left(\frac{\alpha d^2}{6A_n}\right)^{k-1} \\ &\quad \cdot \left(\frac{[4\pi - 3\sqrt{3} + \alpha k]d^2}{6A_n}\right)^{n-1-k} \\ &= G(k,n)\frac{d^{2(n-1)}}{A_n^{n-1}} . \end{aligned} \qquad (5)$$

Combining (5) and (1), we obtain

$$\begin{aligned} Pr\{T_h(n) = k\} &\le \frac{6\pi\alpha^{k-1}[4\pi - 3\sqrt{3} + \alpha k]^{n-1-k} d^{2(n-1)}}{6^{n-1}A_n^{n-1}} \\ &\le \frac{6\pi\alpha^{k-1}[4\pi - 3\sqrt{3} + \alpha k]^{n-1-k} d^{2(n-1)} 6^{n-1}}{6^{n-1}[(6\pi + \alpha(n-2))d^2]^{n-1}} \\ &= \frac{6\pi\alpha^{k-1}[4\pi - 3\sqrt{3} + \alpha k]^{n-1-k}}{(6\pi + \alpha(n-2))^{n-1}} , \end{aligned}$$

proving Theorem 2. □

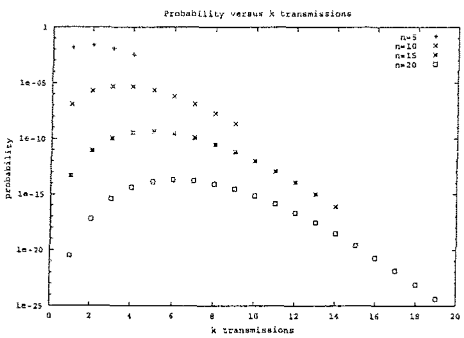

Fig. 2. Probability of number of transmissions for n independent uniformly spaced nodes.

As the above analysis of Theorem 2 and plot of Figure 2 shows, the upper bound probability of successful information dissemination with a fixed number of transmissions is small given a lower bound constraint on the allowable area of uniformly placing nodes. This result justifies providing some form of connectivity topology mapping prior to information dissemination.

III. Comparing Communication Topologies

As an initial study, we propose three classes of topologies and compare the graph properties of these by simulation. These graphs are: (1) minimum radius, (2) relative neighborhood graph and (3) minimum spanning tree. We will discuss each class of graphs in more detail in forthcoming subsections. The graph properties we are interested in are:

- **radius**: proportional to power over data rate.
- **hop diameter**: the number of hops (network diameter) can be used as a lower bound for the number of transmissions.
- **edge density**: fewer edges simplifies the global scheduling.
- **node degree**: affects the number of time slots (or frequencies) needed in local scheduling, i.e., non-interfering nodes can use the same time slots (frequency) for local communication.
- **number of biconnected components**: this shows the number of weak points within the network.
- **size of largest biconnected component**: used to measure the network robustness.

In our simulation, we use sequential algorithms (single processor) to compute the graphs and then compare the graph properties of the produced topologies. For a real application, these algorithms must be redesigned to be distributed* over the nodes.

A. Minimum Radius (minR)

Given a set of randomly placed nodes and assuming that each node must use the same transmission radius, we find the smallest radius d which guarantees network connectivity. Let s denote the side of a square where $s^2 = A$ is the area where the nodes are randomly placed. The algorithm iteratively performs a binary search for the smallest d. In each iteration, the algorithm computes a graph with transmission radius d and checks if the graph is connected. If the graph is connected, then d is decreased, otherwise, d is increased. The algorithm proceeds in iterations until we find the smallest d such that, using d, the communication graph is connected but when using $d-1$, the communication graph is partitioned. The sequential computational cost of this is $O(n^2 \log s)$.

B. Relative Neighborhood Graph

The relative neighborhood graph (RNG) of a node set V in Euclidean space is the graph $G = (V, E)$, where $(v_i, v_j) \in E$ if and only if there is no node $v_z \in V$ such that $\| l_i - l_z \| < \| l_i - l_j \|$ and $\| l_j - l_z \| < \| l_i - l_j \|$, or equivalently, the edge between nodes v_i and v_j is valid if there does not exists any node closer to both v_i and v_j. Referring back to Figure 1, a radius of $\| l_i - l_j \|$ is used for the pair of nodes v_i and v_j. Note that, in RNG, a different radius may be used for each pair of nodes, and so for Figure 1, we could have $d_i \neq d_j$. If the intersection of $A(v_i)$ and $A(v_j)$ does not contain any other nodes, then Node v_i and Node v_j are relative neighbors (i.e. they are directly connected). The RNG is a super-graph of the minimum spanning tree, and it is a sub-graph of the Delaunay triangulation. Supowit[5] presented a sequential algorithm which takes $O(n \log n)$ operations to compute RNG.

C. Minimum Spanning Tree

Since the minimum spanning tree (MST) is a subgraph of RNG, we use RNG in the computation of MST. Note that RNG is a subgraph of the Delaunay triangulation, and the Delaunay triangulation is a planar graph. Thus, the number of edges in RNG is bounded by $3n-6$. We then only need to examine $O(n)$ edges for inclusion/exclusion in the MST. The algorithm first sorts the edges with respect to the edge length, from shortest to longest. This takes $O(n \log n)$ operations. An edge is included in MST if it does not create a cycle in the graph. This

*Since our main interest here is in the graph topologies, we do not consider the distributed computational complexity at this early stage.

is performed by using disjoint sets. Nodes that are connected are placed in the same set. If the tested edge connects two nodes belonging to different sets, then the edge is added to MST and the two sets are unioned. If the tested edge connects two nodes belonging to the same set, then this edge creates a cycle and it is rejected. The number of operations for computing MST is bounded by $O(n)$, given the RNG is pre-computed.

D. Simulation Results

For our simulation† runs, we generated n nodes, randomly placed in an area A, where $5 \leq n \leq 800$, and A is a fixed area of 600^2 units^2, and diagonal $600\sqrt{2}$ units. Two uniformly distributed random integers are generated as the coordinate of each node. For each n, we make 1000 runs. In each run, we use the same set of nodes for the computation of the minimum radius graph (minR), the relative neighborhood graph (RNG) and the minimum spanning tree (MST).

From the simulation, we observe the following:

• **radius**: In minR, every node is required to use the same radius d; thus, d is the smallest radius which renders a connected graph. For RNG and MST, the radius is the longest edge (in Euclidean distance) in the graphs, assuming different transmission radii were possible. Figure 3 shows the plot of the average max-

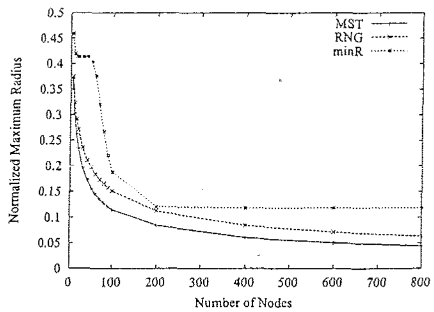

Fig. 3. Maximum transmission radius, averaged over 1000 runs.

imum radius with respect to the number of nodes, normalized with respect to the diameter of A which is $600\sqrt{2}$. As expected, the radius decreases as the number of nodes increases for all three graphs. On the average, MST requires a smaller radius than RNG, where RNG requires a smaller radius than minR. Note that as n increases, the performance of RNG is closer to MST than it is to minR.

• **hop diameter**: the hop diameter of a network is the maximum number of hops among the shortest paths connecting any pair of nodes. This can be used as a lower bound for the number of transmissions required for broadcast. Therefore, it is important to obtain a topology which minimizes the hop diameter. Note that, a partitioned network has hop diameter $+\infty$. Figure 4 shows the plot of hop diameters with respect to the number of nodes. On the average, minR has the lowest hop diameter and MST has the highest. It is worth noting that the RNG hop diameter is closer to the minR than it is to the MST, which means RNG is almost as good as minR in this respect.

† We have implemented the sequential algorithms in JAVA (version 1.2) on a Sun Ultra-10 workstation.

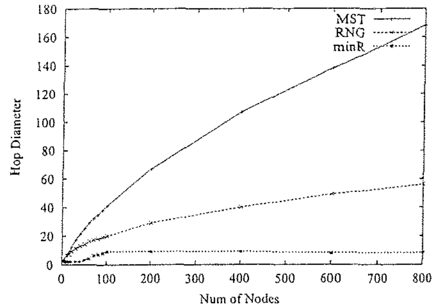

Fig. 4. hop diameter, averaged over 1000 runs.

• **edge density**: the edge density of a graph is computed relative to the maximum number of all possible edges. A graph with n nodes can have at most $\frac{n(n-1)}{2}$ edges. Let this number be $maxE$. The density of a graph $G = (V, E)$ is defined to be $|E|/maxE$, where density is a real number between 0 and 1. We can then compare the densities of MST, RNG and minR. From Figure 5, we observe that the edge densities of both MST

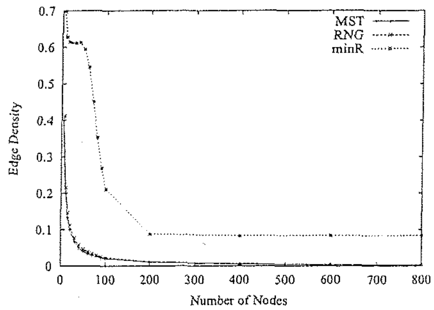

Fig. 5. edge density, normalized and averaged over 1000 runs.

and RNG are very low. As a matter of fact, the plots of MST and RNG almost coincide with each other. In Figure 5, minR has a higher edge density, however, it also decreases very fast as the number of node increases. This is to be expected because as n increases, the number of all possible edges increases quadratically. On the other hand, as n increases, the radius in minR decreases (Figure 3), resulting in fewer edges. Thus, the edge density decreases. A lower edge density may lead to a shorter transmission schedule.

• **node degree**: the node degree is the number of neighbors having direct communication with the node. This affects the scheduling of transmissions. A higher node degree implies that a longer schedule is needed. For each graph, we find the node with the highest node degree, defined as the maximum degree of the graph. In Figure 6, RNG and MST have low node degrees compared to minR. As n increases, the maximum node degree in MST and RNG approaches a small constant. On the other hand, the maximum node degree of minR appears to increase linearly with respect to n. This makes RNG and MST more scalable when local scheduling is used.

• **number of biconnected components**: the number of biconnected components reveals the number of weak points within the network topology. Since biconnected components are connected by articulation points whose failure results in a parti-

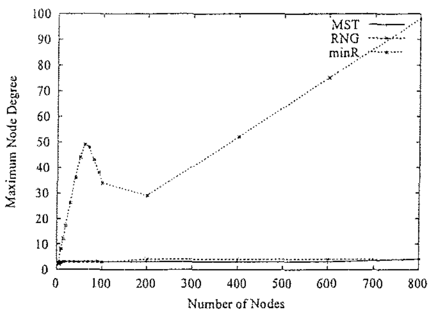

Fig. 6. maximum node degree, averaged over 1000 runs.

tioned network, fewer biconnected components implies a more fault-tolerant network. Since MST is a tree, it does not contain biconnected components (unless we consider each node to be biconnected with itself). Figure 7 shows that minR has fewer biconnected components than RNG. To clarify, in the sparse subgraphs of RNG, each sub-graph may have a tree topology. In that case, each node is counted as a single biconnected component. This may explain why the number of biconnected components in RNG seems to be much higher, compared to minR.

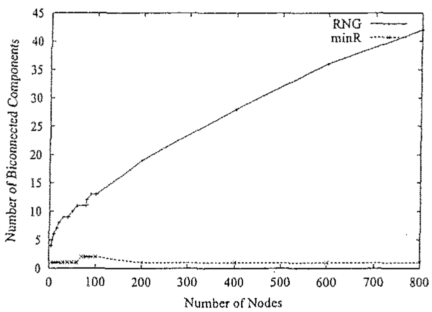

Fig. 7. number of biconnected components, averaged over 1000 runs.

• **largest biconnected component size**: by examining the largest biconnected component, we can determine what percentage of the nodes are not biconnected with the majority of nodes. If the largest biconnected component contains 90% of the nodes, then even if the number of biconnected components is high, we are guaranteed that 90% of the network is fault-tolerant. Figure 8 shows that for most n values, the largest biconnected component in minR contains over 90% of the nodes. The performance of RNG is not far behind with 86% at $n = 100$ and over 90% for $n \geq 200$.

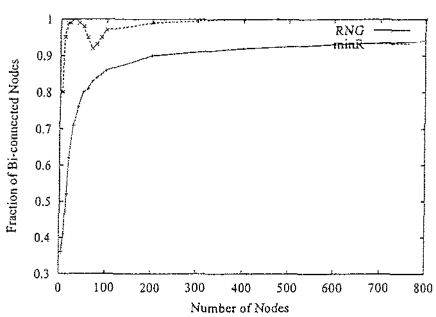

Fig. 8. fraction of nodes in largest biconnected component, normalized and averaged over 1000 runs.

IV. Conclusion

The desire to efficiently reduce the overall energy-per-bit of a node and the analysis indicating the low probability of a bit of information successfully being disseminated among the other nodes with zero knowledge of network topology motivated our study into the graph connectivity for reducing overall transmissions.

From the simulation results, we motivate minR, RNG and MST as follows. To guarantee connectivity, we need a spanning tree at the least. MST is a good choice because it strives to minimize power for fixed data rate. The MST is good in terms of the average maximum transmission radius, edge density and maximum node degree. However, the MST is not fault-tolerant because any node or edge failure will partition the network. It then makes sense to look at a super-graph of MST which still has some of the good graph properties of the MST. For this, we proposed the RNG. The RNG also is good in terms of transmission radius, edge density, and maximum node degree. In addition, our simulation shows that for $n \geq 100$ (or the node density $\geq \frac{n}{A} = \frac{100}{600^2} = \frac{1}{3600}$), the largest biconnected component of RNG contains at least 86% of the nodes. Although this is not as good as minR, it is close. The RNG may have a higher number of biconnected components. However, since RNG's largest biconnected component contains the majority of nodes, this offsets the importance of the number of biconnected components. Concerning hop diameter, RNG is better than MST and worse than minR. However, RNG's hop diameter is closer to minR than it is to MST. The minR is good in terms of the hop diameter, the number of biconnected components and the average largest biconnected component size. However, minR's disadvantages are the higher transmission radius, higher edge density and a node degree which increases linearly with respect to the number of nodes. In light of the above, we suggest that RNG can be a good candidate to consider as a target topology for communication. From the simulation, RNG shows good graph properties when compared with minR and MST.

For future work, we propose to investigate other topologies related to RNG, RNG's implications on amplifiers, and how to compute RNG distributively in an ad-hoc network.

References

[1] Jonathan R. Agre and Loren Clare, "An integrated architecture for cooperative sensing networks," *Computer*, pp. 106–108, May 2000.

[2] Kevin Quirk, Meera Srinivasan, and Jonathan R. Agre, "Cooperative modulation techniques for long haul relay in sensor networks," Nov 2001, To be presented at GLOBECOMM 2001.

[3] Piyush Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," in *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming. Edited by W.M.McEneany, G. Yin, and Q. Zhang*, pp. 547–566. Birkhauser, Boston, 1998.

[4] Juraj Hromkovič, Ralf Klasing, Burkhard Monien, and Regine Peine, "Dissemination of information in interconnection networks," in *Combinatorial Network Theory. Edited by D.Z. Du and F. Hsu*, pp. 125–212. Kluwer Academic Publisher, 1995.

[5] Kenneth J. Supowit, "The relative neighborhood graph, with an application to minimum spanning trees," *Journal of the ACM*, vol. 30, no. 3, pp. 428–448, 1983.